Weekly report

1 Done

1.1 Vast Modification

- Removed the reference errors mentioned in the comments of the second review.
- Re-designed figures to shorten the length.
- Added author information and acknowledgements

1.2 DP4G (Differential Privacy for Graphs)

Now we have some questions that need to be solved.

- How to facilitate users to define the parameter ϵ? Existing studies proposed various formula derivations from the perspective of the concept rather than user needs.
- How to explain uncertainty? On the one hand, uncertainty depends on adversary/analyst's background knowledge, which we can't predict. On the other hand, related computation should meet the need of interaction. However, some of them are with a high time complexity.

For both of the questions, I plan to consult Prof. Ji.

1.3 Reading

- Technical Privacy Metrics: a Systematic Survey. [CSUR2018]

  This is a comprehensive survey paper on privacy preservation models. It summarizes privacy domains, privacy metrics and future research directions. The following metrics can be applied in our study:

  1) Uncertainty: This work only mentioned the uncertainty of adversary's estimate (adversary's expected estimation). Actually, we can also discuss the uncertainty of analyst's inferences. The two kinds of uncertainty are different from prior information (background knowledge) and posterior information (conclusions). I consider this metrics as the most significant role in adjusting the size of noise. However, it is difficult to understand it by a group of numbers (like anonymity set size, entropy, etc.). Visual interpretation is necessary here.
  2) Parameter of differential privacy (epsilon): In statistical databases, differential privacy guarantees that any disclosure is equally likely (within a small multiplicative factor ϵ) regardless of whether or not an item is in the database. The author presented a conclusion: the choice of the parameter ϵ is difficult: values reported in the literature vary from 0.01 to 100. That is reason why we need to explain the model to users and allow users to set parameter interactively.

- Privacy models for big data: a survey. [IJBDI2016]

The major challenges in the case of privacy preservation of social network data:

1) It is very difficult to model the background knowledge.
   Without background knowledge, we can't define adversary's uncertainty definitely. We need fix this issue as well.
2) It is quite challenging to find the information loss in social network data.
3) Finding out the best technique for anonymizing social network data is yet another interesting challenge.

- A Random Matrix Approach to Differential Privacy and Structure Preserved Social Network Graph Publishing. [??2013]

This is a differential privacy approach for graph publishing.

Algorithm 1: $\widehat{A} = \texttt{Publish}(A, m, \sigma^2)$

Input: (1) symmetric adjacency matrix $A \in \mathbb{R}^{n \times n}$
(2) the number of random projections $m < n$
(3) variance for random noise σ^2

Output: $\widehat{A}$

1 Compute a random projection matrix P, with $P_{i,j} \sim \mathcal{N}(0, 1/m)$
2 Compute a random perturbation matrix Q, with $Q_{i,j} \sim \mathcal{N}(0, \sigma^2)$
3 Compute the projected matrix $A_p = AP$
4 Compute the randomly perturbed matrix $\widehat{A} = A_p + Q$

The projections are achieved by spectral clustering depending on eigenvectors of the adjacency matrix.

Algorithm 2: Spectral Clustering

Input: (1) Adjacency Matrix $A \in \mathbb{R}^{n \times n}$
(2) Number of clusters k

Output: Clusters $C_1, ..., C_k$

1 Compute first k eigenvectors $\mathbf{u}_1, .., \mathbf{u}_k$ of A
2 Get matrix $U \in \mathbb{R}^{n \times k}$ where ith column of U is $\mathbf{u}_i$
3 Obtain clusters by applying $k-$means clustering on matrix U

The eigenvector can be counted as a kind of "skeleton".

- Social Graph Publishing with Privacy Guarantees. [ICDCS 2016]

Another graph publishing approach based on random matrix.

- TopoText: Context-Preserving Text Data Exploration Across

Multiple Spatial Scales [CHI2018]

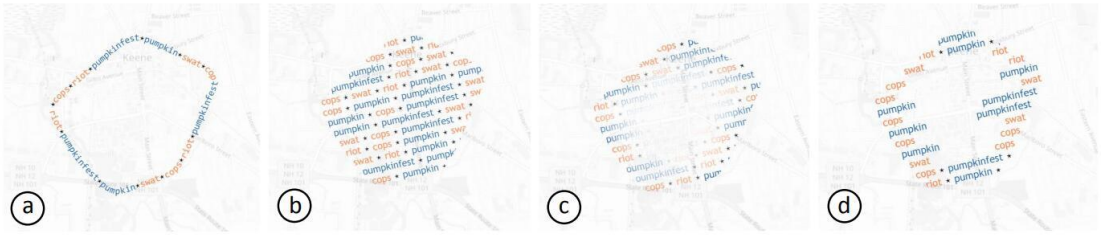

Figure 4. Design alternatives for visualizing the text data on a single aggregate. (a): The text labels are placed along the boundary; (b): The text labels are filled within the area of the aggregate; (c): The space-filling visualization is enhanced by applying a transparency gradient on the text labels; (d): The text labels that are close to the boundary are placed inside the aggregate.

They are four interesting designs for text data on multiple spatial scales. The hierarchy effects are shown as below.

1.4 Preparing for going abroad

2 Progress

Item	Deadline	Current progress	Remark
VAST R2 and preview video	8.1	-	
Courseware revision	9.1	Sent the current version by email.	
Go abroad	11.18	Handling ds-2019	
Privacy program	10.31	Surveying.	I will focus on it next week.